Data Preprocessing: Final Project

Sentiment Analysis in E-commerce Platform (Amazon)

Problem Statement: The e-commerce industry, especially on large platforms like Amazon, generates a massive amount of customer reviews daily. These reviews offer valuable insights into customer satisfaction and product performance. However, manually analyzing such vast quantities of textual data is both labor-intensive and time-consuming. Sentiment analysis of Amazon customer reviews aims to address this challenge by providing an automated approach to evaluate customer feedback.

Significance: Sentiment analysis of customer reviews on e-commerce platforms like Amazon has numerous practical applications. It can help e-commerce platforms:

- Enhance product recommendations by understanding customer preferences.
- Identify product quality issues and areas for improvement.
- Monitor customer satisfaction trends to improve customer service.
- Automatically categorize and prioritize reviews for immediate response and action.

Project Goal: The goal of this project is to propose, develop, and evaluate a data-driven model for sentiment analysis of Amazon customer reviews using machine learning and natural language processing techniques.

Dataset:

https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews/data

Data Preprocessing: Final Project

Cough Sound-Based COVID-19 Detection

Problem Statement: COVID-19, caused by the SARS-CoV-2 virus, has had a profound global impact. Timely and accurate detection of COVID-19 cases is crucial for controlling the spread of the virus. Recent research has shown promise in using cough sound analysis as a potential non-invasive method for early detection of COVID-19 infections. This project aims to leverage machine learning and data-centric AI techniques to develop an effective and efficient COVID-19 detection system based on the analysis of cough sounds.

Significance: A cough sound-based COVID-19 detection system can have a significant impact on the early identification and management of COVID-19 cases. It offers a non-invasive and potentially cost-effective method for large-scale screening and monitoring.

Project Goal: The primary goal of this project is to propose, build, and evaluate a data-driven model that can detect COVID-19 infections using cough sound data. This model will analyze the acoustic features of cough sounds and classify them as COVID-19 positive or negative.

Dataset:

Coswara dataset
https://github.com/iiscleap/Coswara-Data/tree/master

Coughvid dataset
https://zenodo.org/records/7024894

Data Preprocessing: Final Project

Lung Cancer Classification

Problem Statement: Lung cancer remains one of the leading causes of cancer-related deaths worldwide. Early and accurate diagnosis is essential for improving patient survival rates, but traditional diagnostic methods can be invasive, time-consuming, and costly. Recent advances in medical imaging and machine learning provide a promising opportunity to develop automated systems for lung cancer classification, enabling faster and more precise diagnosis from medical scans.

Significance: Lung cancer classification has significant implications in the field of medical diagnosis and treatment. An automated classification system can:

- Aid radiologists in identifying malignant and benign lung nodules with greater accuracy.
- Reduce the time required for diagnosis, allowing for quicker treatment decisions.
- Help lower healthcare costs by reducing unnecessary invasive procedures.
- Contribute to early detection, which is crucial for improving patient outcomes.

Project Goal: This project aims to propose, develop, and evaluate a Data-driven model for classifying lung cancer from medical images. The model will be trained to distinguish between cancerous and non-cancerous tissues, aiding medical professionals in diagnosing lung cancer with greater efficiency and accuracy.

Dataset:

https://luna16.grand-challenge.org/Download/

or:

https://www.kaggle.com/datasets/fanbyprinciple/luna-lung-cancer-dataset

Data Preprocessing: Final Project

Fake News Detection System

Problem Statement: The rapid spread of fake news across social media and online platforms poses a significant threat to public discourse and informed decision-making. Misinformation can shape opinions, influence elections, and even incite panic during crises. Manually verifying the authenticity of news articles is a daunting and often impractical task due to the sheer volume of content produced daily.

Significance: The detection of fake news has several important applications:

- Enhancing Credibility: Helps media organizations and platforms maintain credibility by reducing the dissemination of false information.
- Informed Citizenry: Empowers individuals to make better-informed decisions based on reliable information.
- Crisis Mitigation: Reduces the spread of harmful misinformation during critical events, such as public health emergencies or elections.
- Supporting Journalistic Integrity: Aids journalists in identifying misleading information and focusing on credible sources**.**

Project Goal: The goal of this project is to propose, develop, and evaluate a Data-driven model for fake news detection. The model will analyze news articles and social media posts to classify them as real or fake.

Dataset:

https://www.kaggle.com/datasets/saurabhshahane/fake-news-classification